

Combining Flux and Energy Balance Analysis to Model Genome-Scale Biochemical Networks

William J. Heuett (University of Colorado, Boulder) and Hong Qian (University of Washington, Seattle)

Abstract

Stoichiometric Network Theory is a constraints-based, optimization approach for quantitative analysis of the phenotypes of large-scale biochemical networks that avoids the use of detailed kinetics. This approach uses the reaction stoichiometric matrix in conjunction with constraints provided by flux balance and energy balance to guarantee mass conserved and thermodynamically allowable predictions, respectively. However, to date, the flux and energy balance constraints have not been implemented simultaneously because optimization under the combined constraints is nonlinear. We introduce a sequential quadratic programming algorithm that solves the nonlinear optimization problem. The system of fermentation in *Saccharomyces cerevisiae* is used to illustrate the new method. The algorithm allows the use of nonlinear objective functions. As a result, we suggest a novel optimization with respect to the heat dissipation rate of a system. We also emphasize the importance of incorporating interactions between a model network and its surroundings.

Introduction

Genome-wide sequencing efforts and subsequent bioinformatic analysis have provided a collection of reconstructed biochemical networks for a number of living organisms (Edwards and Palsson 2000). As this collection continues to grow, methods for analyzing the cellular functions of the networks become increasingly important. Many mathematical methods have been developed to provide a description of these networks; however, most of these methods rely on rate equations that require detailed, kinetic-rate information. Unfortunately, rate constants are not generally known; nor can they be determined with current experimental techniques *in vivo* or from the genome sequences (Edwards and Palsson 2000).

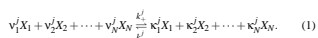
To address the lack of kinetic-rate information, genome-scale, constraints-based models have been developed to describe the functional states, or phenotypes, of many organisms (Westerhoff and Palsson 2004). These optimization approaches are used to analyze biochemical reaction systems in nonequilibrium steady-state (NESS) via constraints, provided by Stoichiometric Network Theory (SNT), that are based on the stoichiometry of the system, i.e., the static, algebraic, topological structure of the biological network that provides the framework within which chemical "motion" must take place (Erdi and Tóth 1989). Using Flux Balance Analysis (FBA) (Edwards and Palsson 2000) and Energy Balance Analysis (EBA) (Beard et al. 2002; Qian et al. 2003), feasible solutions of the optimization problem are required to satisfy mass and energy conservation, as well as the second law of thermodynamics.

FBA has been used to analyze biochemical networks of many organisms including mutant strains of *E. coli* (Edwards and Palsson 2000), *Haemophilus influenza* (Edwards and Palsson 1999), and mitochondrial energy metabolism (Ramakrishna et al. 2001). However, it has since been shown that FBA alone does not guarantee thermodynamically feasible solutions, i.e., reaction fluxes could satisfy mass balance but flow against energy gradients. For this reason, EBA was developed to make the results of the SNT constraints-based approaches more physically realistic while revealing further insight into the mechanisms of these complex biochemical networks (Beard et al. 2002; Qian et al. 2003).

Although algorithms for solving optimization problems subject to FBA or EBA constraints separately are available, an algorithm that explicitly incorporates both types of constraints simultaneously, thus treating the flux and energy on equal footing, does not exist. This is due to the fact that the combination of these constraints creates a nonlinear, constrained optimization problem, which is more difficult to solve. Furthermore, previous studies have only considered linear objective functions when modeling biochemical networks. For example, when studying *E. coli* metabolism, the maximization of the biomass needed to reproduce has been used (Edwards and Palsson 2000; Beard et al. 2002). Arguments based on experimental evidence can be made regarding what objective function is believed to be appropriate for each particular system. However, greater insight can be obtained if *in silico* experiments can be done to test different, possibly nonlinear, objective functions for each system. We present an algorithm that incorporates FBA and EBA constraints simultaneously and allows the testing of nonlinear objective functions.

Stoichiometric Network Theory

To derive the constraints used in the optimization methods, we begin by considering a biochemical reaction network with M reactions and N chemical species denoted by X_i , $i = 1, 2, \dots, N$, where the j^{th} reaction can be represented by the generic reaction



If this system is being driven by external flux interactions with its surroundings, the kinetics equations used to model the dynamics of such a reaction network are given in the compact form

$$\frac{d\mathbf{x}}{dt} = \mathbf{S}\mathbf{J} + \mathbf{J}^{\text{ext}}, \quad (2)$$

where \mathbf{x} represents the N -dimensional vector of species concentrations, \mathbf{J} represents the M -dimensional vector of reaction fluxes, \mathbf{J}^{ext} is the N -

dimensional vector of external fluxes, and $\mathbf{S} \in \mathbb{R}^{N \times M}$ is known as the stoichiometric matrix with i^{th} row and j^{th} column entry given by $(S)_{i,j} = \kappa_i^j - \nu_i^j$.

The stoichiometric matrix is known once a biochemical network has been reconstructed from the genome sequencing and the bioinformatic analysis has been done. Unfortunately, this is not the case for the reaction rate constants, k_i^j and k_r^j . Rather, the rate constants are not generally known, nor can they be readily measured *in vivo*. Therefore, we must turn to methods of analysis that do not require knowledge of these constants. By writing the kinetics equations as in (2), the unknowns are absorbed into the flux vector \mathbf{J} , making it a vector of unknowns. In any way, this is convenient because we do not have to be concerned about the form of the kinetics equations, i.e., whether we should use mass action kinetics or Michaelis-Menten kinetics equations, since all that information has also been absorbed into the vector \mathbf{J} .

A biochemical network that is interacting with its surroundings by constantly being driven by external fluxes flowing into and out of it will go to a NESS (Hill 1974; Qian and Beard 2005). Therefore, a lot of insight about the system's properties can be gained by studying it in NESS. This is the basis of SNT where the stoichiometric structure of the biochemical network is used to describe a system's phenotypic functions.

Flux Balance Analysis

When a system is in NESS, the concentrations of the chemical species are not changing but there is a nonzero flux distribution flowing through the network. From (2), we see that the NESS flux distribution must satisfy

$$\mathbf{S}\mathbf{J} = -\mathbf{J}^{\text{ext}}, \quad (3)$$

which is known as the flux balance constraint of FBA (Edwards and Palsson 2000). This constraint is similar to Kirchhoff's current law of electrical circuit theory and says that the NESS flux distribution through the biochemical network must conserve mass.

Additional upper and lower bounds can be applied to the NESS fluxes. In most cases, the bounds for the internal fluxes are infinite, but finite bounds may be imposed if there is experimental evidence that suggests such constraints are physically relevant. Bounds may also be applied to test how the flux distribution through the network adjusts if the enzyme that drives a particular reaction is absent or limited due to a genetic mutation or disease. Imposing bounds on the external fluxes is useful for defining the medium that the living system is subject to as well as the biomass and waste being removed from the system.

The FBA constraints require the NESS fluxes to conserve mass, but this does not guarantee that the fluxes are thermodynamically feasible. According to the laws of thermodynamics, the flux of any given reaction must flow down the reaction's potential gradient (Qian and Beard 2005). A flux flowing against an energy gradient would constitute a perpetual motion machine, able to create energy from nothing, which is not possible. Therefore, the laws of thermodynamics must be incorporated to be certain that the feasible fluxes are restricted to thermodynamically feasible fluxes.

Energy Balance Analysis

EBA is the theory and methodology for enforcing the laws of thermodynamics in the SNT approaches (Beard et al. 2002; Qian et al. 2003). Just as the constraints of FBA result from the structure of the stoichiometric matrix \mathbf{S} , so do the constraints of EBA. If \mathbf{r} is defined as the N -dimensional vector of chemical potentials, the M -dimensional vector of reaction potentials, $\Delta\mu$, is given by $\mathbf{S}^T \mathbf{r} = \Delta\mu$. Assuming \mathbf{S} has l linearly independent rows, $r \in \mathbb{R}^N$, the nullspace matrix $\mathbf{K} \in \mathbb{R}^{N \times (M-l)}$ can be constructed with columns that form a basis for the nullspace of \mathbf{S} , so that $\mathbf{S}\mathbf{K} = \mathbf{0}$. The nullspace matrix describes the internal loops of the biochemical network and can be used to give a constraint requiring the conservation of energy, i.e., the first law of thermodynamics, as

$$\mathbf{K}^T \mathbf{S}^T \boldsymbol{\mu} = \mathbf{K}^T \Delta\boldsymbol{\mu} = \mathbf{0}. \quad (4)$$

This constraint requires that the sum of reaction potentials around any cycle of reactions equals 0, which is similar to Kirchhoff's voltage or loop law of electrical circuit theory, and is known as the energy balance constraint of EBA (Beard et al. 2002; Qian et al. 2003).

By defining the nonnegative forward and reverse reaction fluxes, \mathbf{J}_+ and \mathbf{J}_- , respectively, with j^{th} entries representing the one-way fluxes through the j^{th} reaction, the net flux distribution through the reactions of the network is given by $\mathbf{J} = \mathbf{J}_+ - \mathbf{J}_-$. Similarly, the j^{th} reaction potential is given by (Qian and Beard 2005)

$$\Delta\mu^j = RT \ln \left(\frac{p_+^j}{p_-^j} \right), \quad (5)$$

where R is the gas constant and T is the temperature. This relationship leads directly to the second law of thermodynamics, i.e.,

$$-j^{\text{th}} \Delta\mu^j = -RT \left(p_+^j - p_-^j \right) \ln \left(\frac{p_+^j}{p_-^j} \right) \geq 0, \quad (6)$$

which says that the system must dissipate heat and entropy must increase as a result of the work being done on the system through the external fluxes. The inequality in (6) is an equality if and only if $j^{\text{th}} = \Delta\mu^j = 0$, in which case the reaction would be in equilibrium. The total heat dissipation rate of the living system is given by $hdr = -\mathbf{J}^T \Delta\boldsymbol{\mu} > 0$. It is mathematically possible to have $hdr \rightarrow 0$ in the limit as $\mathbf{J} \Delta\boldsymbol{\mu} \rightarrow \mathbf{1}$ component-wise while maintaining $\mathbf{J} = \mathbf{J}_+ - \mathbf{J}_-$. However, to prevent this physically unrealistic possibility, we propose the additional constraint

$$(hdr)_{lb} \leq hdr \leq (hdr)_{ub}. \quad (7)$$

The Optimization Problem

We would like to describe the phenotypes of a living organism in NESS by simultaneously imposing the FBA and EBA constraints and minimizing a given smooth, linear or nonlinear, objective function. That is, we want to solve the nonlinear, constrained-optimization problem represented by

$$\begin{aligned} \min_{\mathbf{J}, \mathbf{J}_+, \mathbf{J}_-} \quad & f(\mathbf{J}, \mathbf{J}_+, \mathbf{J}_-, \mathbf{J}^{\text{ext}}, \Delta\boldsymbol{\mu}) \\ \text{s.t.} \quad & \mathbf{S}\mathbf{J} + \mathbf{J}^{\text{ext}} = \mathbf{0} \\ & \mathbf{K}^T \Delta\boldsymbol{\mu} = \mathbf{0} \\ & \text{diag} \left(e^{(\Delta\mu^j/RT)} \right) \mathbf{J}_- - \mathbf{J}_+ = \mathbf{0} \\ & \mathbf{J} = \mathbf{J}_+ - \mathbf{J}_- = \mathbf{0} \\ & \mathbf{J}_{lb} \leq \mathbf{J} \leq \mathbf{J}_{ub} \\ & \mathbf{0} \leq \mathbf{J}_+ < \infty \\ & \mathbf{0} \leq \mathbf{J}_- < \infty \\ & \mathbf{J}_{lb}^{\text{ext}} \leq \mathbf{J}^{\text{ext}} \leq \mathbf{J}_{ub}^{\text{ext}} \\ & \Delta\mu_{lb}^j \leq \Delta\mu^j \leq \Delta\mu_{ub}^j \\ & (hdr)_{lb} \leq hdr \leq (hdr)_{ub} \end{aligned} \quad (8)$$

where the boundary constraints are meant to be satisfied component-wise. The objective function is assumed to have a biological meaning. It should be considered as a formal, quantitative hypothesis about the network of interest. The optimization generates predictions, which have to be tested, for the hypotheses.

Let n be the total number of constraints in the problem. Then, if $\mathbf{x} = (\mathbf{J}^T, (\mathbf{J}^{\text{ext}})^T, \mathbf{J}_+^T, \mathbf{J}_-^T, \Delta\boldsymbol{\mu}^T)^T$, the Lagrangian of the nonlinear optimization problem (8) is

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \sum_{i=1}^n \lambda_i c_i(\mathbf{x}), \quad (9)$$

where $\boldsymbol{\lambda}$ is the n -dimensional vector of Lagrange multipliers, which are required to satisfy strict complementarity conditions, and $c_i(\mathbf{x})$ is the i^{th} constraint evaluated at \mathbf{x} . If \mathbf{x}^* is a local solution of (8) and the set $\{\nabla c_i(\mathbf{x}^*) | c_i(\mathbf{x}^*) = 0\}$ is linearly independent, a first-order, necessary, optimality conditions require that there exists a vector $\boldsymbol{\lambda}^*$ such that $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}$.

An optimization problem in the form of (8) is well suited to be solved using a Sequential Quadratic Programming (SQP) algorithm (Nocedal and Wright 1999). The basic idea of an SQP method is to iteratively step toward an optimal solution, i.e., a root of $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$, by approximating the original nonlinear problem at each step by a quadratic subproblem. A simple interpretation of an SQP algorithm is to view it as an application of Newton's method to the KKT optimality conditions (Nocedal and Wright 1999). It is straightforward to establish local convergence of the SQP algorithm, yielding optimal solutions that are both mass balanced and thermodynamically feasible. To obtain the results presented here, the SQP algorithm was implemented using the MATLAB (Mathworks, Natick, Massachusetts, United States) computing environment and its built in functions, including its quadratic programming algorithm *quadprog*.

Saccharomyces cerevisiae Glycolysis

We now focus on the well studied system of glycolysis in yeast, in particular *Saccharomyces cerevisiae*. Yeast have the ability to metabolize natural sugars and remain dispersed in a medium until the substrate has been metabolized, at which point the cells flocculate and settle out, leaving a clear liquid above the yeast (McKinney 2004). This process produces alcohol and is the reason yeast is so valuable in the fermentation industry. Overall, the net reaction of fermentation is the conversion of glucose to ethanol and carbon dioxide

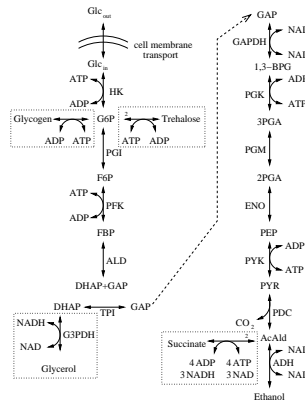
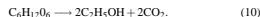


Figure 1. Illustration of yeast glycolysis.

Table 1. Flux boundaries ($\mu\text{mol} \cdot \text{min}^{-1} \cdot \text{mg protein}^{-1}$) for yeast glycolysis (Teusink et al. 2000).

rxn	transport	HK	PGI	PK	ALD	TPI	GAPDH
J_{ub}	—	—	—	—	—	—	-24.3
J_{lb}	0.36	0.84	1.26	0.68	1.19	8.4	4.4
J_{ub}	PGK	PGM	ENO	PVK	PDC	ADH	—
J_{lb}	-4.8	—	—	—	—	-3.0	—
J_{ub}	—	9.4	1.35	0.05	0.65	—	—

Under anaerobic conditions, most of the energy from the sugar is transferred to ethanol production while growth of the yeast cells is minimized (McKinney 2004). An FBA constrained optimization of the unbranched network, i.e., excluding the reactions in dotted boxes in Figure 1, to maximize ethanol output subject to the flux bounds listed in Table 1 yields the optimal flux distribution shown in Table 2. These fluxes are consistent with results from Teusink et al. (2000), who developed a Michaelis-Menten kinetics model of yeast glycolysis. It shows that the unbranched network is limited by the maximum flux of PDC. This failure of the "lower" part of glycolysis to keep up with the flux through the "upper" part is typical of the phenotype observed in Tps1 mutants, a gene associated with trehalose-6-phosphate synthase, the enzyme responsible for trehalose synthesis (Bonini et al. 2000; Teusink et al. 2000).

A very different story can be told, however, if the branches of the network are taken into account. Using the constant output fluxes experimentally determined by Teusink et al. (2000) for the 4 boxed branches in Figure 1, an FBA constrained optimization of the branched system illustrates that the PDC limitation is relieved by the additional external fluxes. The resulting optimal flux distribution is shown in Table 2. This shift in fluxes through the network due to the incorporation of additional structural information about the system emphasizes the importance of understanding how simplified subset networks being modeled interact with other cellular functions. Incorporating such information makes a model more physically relevant and can improve accuracy of results. Furthermore, these links will play a crucial role when models are combined in an effort to create a unified model of a living organism.

Another important factor to consider is the heat produced by fermentation. If not removed, it can be lethal or may cause the yeasts to pasteurize themselves, halting fermentation (Ameringer and Singleton 1977). Under anaerobic conditions with a complex medium and glucose as the substrate, a continuous culture of *S. cerevisiae* has a specific rate of heat production (or hdr) of $0.2 \text{ W} \cdot \text{g}^{-1}$ (Cortassa et al. 2002). Table 2 lists the optimal solutions found for the unbranched and branched yeast glycolysis systems by using the SQP algorithm to minimize

$$f(\mathbf{J}^{\text{ADH}}, \Delta\boldsymbol{\mu}) = -\mathbf{J}^{\text{ADH}} \cdot \frac{\Delta\mu^{\text{ADH}}}{2} \quad (11)$$

subject to the FBA, EBA, and heat dissipation constraints. Again, we see that the unbranched system is limited by the maximum PDC reaction rate. However, inclusion of the branches to other subnetworks of yeast metabolism provides relief from this limitation. As a result, the maximum amount of glucose uptake into the cell is allowed and a realistic estimate of the maximum rate of ethanol production is obtained. This rate of ethanol production, $0.51 \mu\text{mol} \cdot \text{min}^{-1} \cdot \text{mg protein}^{-1}$, is consistent with the experimentally measured rate from Teusink et al. (2000). For both the unbranched and branched systems, the optimal solution corresponds to a $hdr = 0.2 \text{ W} \cdot \text{g}^{-1}$, increasing confidence that the optimal solutions obtained are physically relevant to the extent that the homeostatic and thermodynamic information is accurate.

Table 2. Optimal solutions for yeast glycolysis.

rxn	Unbranched System				Branched System			
	FBA	FBA and EBA	FBA	FBA and EBA	FBA	FBA and EBA	FBA	FBA and EBA
J_{ub}	0.33	0.33	1.02	0.69	-971	0.36	0.36	0.75
HK	0.33	0.33	1.02	0.69	-971	0.36	0.36	0.75
PGI	0.33	0.33	1.02	0.69	-971	0.30	0.30	0.70
PK	0.33	0.33	1.02	0.69	-971	0.30	0.30	0.70
ALD	0.33	0.33	1.02	0.69	-971	0.30	0.30	0.70
TPI	0.33	0.33	1.02	0.69	-971	0.23	0.23	0.64
GAPDH	0.65	0.65	1.21	0.56	-1943	0.54	0.53	0.87
PGK	0.65	0.65	1.21	0.56	-1943	0.54	0.53	0.87
PGM	0.65	0.65	1.21	0.56	-1943	0.54	0.53	0.87
ENO	0.65	0.65	1.21	0.56	-1943	0.54	0.53	0.87
PVK	0.65	0.65	1.21	0.56	-1943	0.54	0.53	0.87
PDC	0.65	0.65	1.21	0.56	-1943	0.54	0.53	0.87
ADH	0.65	0.65	1.21	0.56	-1943	0.51	0.51	0.85
ATPase	0.65	0.65	1.21	0.56	-1943	0.32	0.31	0.72
Glycogen	—	—	—	—	—	0.02	0.02	0.23
Trehalose	—	—	—	—	—	0.02	0.02	0.23
Glycerol	—	—	—	—	—	0.07	0.07	0.41
Succinate	—	—	—	—	—	0.01	0.01	0.22

Conclusions

Biochemical reaction networks are effectively being reconstructed by genomic sequencing and bioinformatic analysis. This is done with the hope that it will be useful in describing the cellular functions and phenotypes of living organisms on the genome-scale. The SNT constraints-based optimization approaches take advantage of this information to analyze the systems in NESS without requiring detailed kinetics information. As a result, this approach has been shown to be an accurate, useful tool for studying mutant and disease affected organisms.

When FBA and EBA are combined, they yield results that are both mass balanced and thermodynamically feasible. The addition of the heat dissipation constraint adds confidence and physical relevance to the results. More constraints, such as concentration constraints, and experimentally determined information can be incorporated to further restrict the feasible space over which the optimization is done, and yield more insight into the biological processes. There are several other directions for future work to follow, from improvements in the speed of the algorithm to the incorporation of noise effects using stochastic optimization techniques.

The SQP algorithm presented here is a general method for solving nonlinear programming problems. It is a tool that researchers from various different backgrounds can use to quickly perform *in silico* experiments on interesting biological scenarios instead of having to perform several expensive laboratory experiments. Surely, this methodology will continue to play an important role in the future of systems biology.

Acknowledgments

The work of W.J.H. was supported, while a graduate student, by a NSF VIGRE fellowship and a DoD NDSG fellowship. H.Q. was supported by NIH Grant Nos. GM068610 and HL072011. We thank James Burke, Daniel Beard, and Mark Kot for their helpful discussions and input.

REFERENCES

- Ameringer, M. A., & Singleton, V. L. (Eds.) 1977. Wine: An Introduction (2nd ed.). Berkeley: University of California Press.
- Beard, D. A., Liang, S. D., & Qian, H. 2002. Biophys. J. 83, 79.
- Bonini, B. M., et al. 2000. Biochem. J. 350, 261.
- Cortassa, S., Azzi, M. A., Iglesias, A. A., & Lloyd, D. (Eds.) 2002. An Introduction to Metabolic and Cellular Engineering. River Edge: World Scientific.
- Edwards, J. S., & Palsson, B. O. 1999. J. Biol. Chem. 274, 17410.
- Edwards, J. S., & Palsson, B. O. 2000. Proc. Natl. Acad. Sci. USA, 97, 5528.
- Erdi, P., & Tóth, J. 1989. Mathematical Methods of Chemical Reactions: Theory and Applications of Deterministic and Stochastic Models. Manchester: Manchester Univ. Press.
- Hill, T. L. 1974. Free Energy Transduction in Biology: The Steady-State Kinetic and Thermodynamic Formalism. New York: Academic Press.
- McKinney, R. E. 2004. Environmental Pollution Control Microbiology. New York: Marcel Dekker, Inc.
- Nocedal, J., & Wright, S. J. 1999. Numerical Optimization. New York: Springer.
- Qian, H., & Beard, D. A. 2005. Biophys. Chem. 114, 213.
- Qian, H., Beard, D. A., & Liang, S. D. 2003. Eur. J. Biochem., 270, 415.
- Ramakrishna, R., Edwards, J. S., McCulloch, A., & Palsson, B. O. 2001. Am. J. Physiol. Reg. Int. Comp. Physiol., 280, R695.
- Teusink, B. et al. 2000. Eur. J. Biochem., 267, 531.
- Westerhoff, H. V., & Palsson, B. O. 2004. Nature Biotech., 22, 1249.

This miniposter was prepared with Brian Wolven's Poster L^AT_EX macros v2.1.